



PREPARING FOR ADVANCED LIGO: A STAR–GALAXY SEPARATION CATALOG FOR THE PALOMAR TRANSIENT FACTORY

A. A. MILLER^{1,2,6}, M. K. KULKARNI^{2,3}, Y. CAO², R. R. LAHER⁴, F. J. MASCI⁵, and J. A. SURACE⁴

¹ Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive,

MS 169-506, Pasadena, CA 91109, USA; amiller@astro.caltech.edu

² California Institute of Technology, Pasadena, CA 91125, USA

³ University of California–Berkeley, Berkeley, CA 94720, USA

⁴ Spitzer Science Center, California Institute of Technology, Pasadena, CA 91125, USA

⁵ Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA

Received 2016 August 30; revised 2016 November 16; accepted 2016 December 2; published 2017 January 16

ABSTRACT

The search for fast optical transients, such as the expected electromagnetic counterparts to binary neutron star mergers, is riddled with false positives (FPs) ranging from asteroids to stellar flares. While moving objects are readily rejected via image pairs separated by ~ 1 hr, stellar flares represent a challenging foreground, significantly outnumbering rapidly evolving explosions. Identifying stellar sources close to and fainter than the transient detection limit can eliminate these FPs. Here, we present a method to reliably identify stars in deep co-adds of Palomar Transient Factory (PTF) imaging. Our machine-learning methodology utilizes the random forest (RF) algorithm, which is trained using $>3 \times 10^6$ sources with Sloan Digital Sky Survey (SDSS) spectra. When evaluated on an independent test set, the PTF RF model outperforms the SExtractor star classifier by $\sim 4\%$. For faint sources ($r' \geq 21$ mag), which dominate the field population, the PTF RF model produces a $\sim 19\%$ improvement over SExtractor. To avoid false negatives in the PTF transient-candidate stream, we adopt a conservative stellar classification threshold, corresponding to a galaxy misclassification rate of 0.005. Ultimately, $\sim 1.70 \times 10^8$ objects are included in our PTF point-source catalog, of which only $\sim 10^6$ are expected to be galaxies. We demonstrate that the PTF RF catalog reveals transients that otherwise would have been missed. To leverage its superior image quality, we additionally create an SDSS point-source catalog, which is also tuned to have a galaxy misclassification rate of 0.005. These catalogs have been incorporated into the PTF real-time pipelines to automatically reject stellar sources as non-extragalactic transients.

Key words: catalogs – galaxies: statistics – methods: data analysis – methods: statistical – stars: statistics – surveys

1. INTRODUCTION

The classification or separation of stars versus galaxies in astronomical images is an old problem with many important modern applications. At a very basic level, number counts of bright galaxies as a function of magnitude show that the universe does not have a homogeneous “Euclidean” geometry (Yasuda et al. 2001). More importantly, the accurate separation of stars and galaxies in faint samples significantly improves our ability to (i) measure galaxy–galaxy correlation functions (e.g., Ross et al. 2011; Ho et al. 2015), (ii) map the signature of baryon acoustic oscillations (Anderson et al. 2014), (iii) search for dwarf galaxies by looking for stellar overdensities (e.g., Belokurov et al. 2007), (iv) detect the weak-lensing signal from cosmic shear (Soumagnac et al. 2015 and references therein), and (v) trace structure in the Milky Way halo (e.g., Belokurov et al. 2006; Jurić et al. 2008), among other things.

The array of scientific problems dependent on star–galaxy separation is disparate, meaning that the construction of any such catalog should be application specific. For time-domain surveys aiming to identify transients, a reliable star–galaxy catalog immediately informs researchers of the galactic or extragalactic origin of newly discovered candidates.

The Palomar Transient Factory (PTF; Rau et al. 2009; Law et al. 2009) is a dedicated survey of the variable sky utilizing the CFH12K mosaic camera on the Palomar 48-inch telescope (P48). The initial phase of this experiment ended in 2012, while

the current iteration, the intermediate Palomar Transient Factory (iPTF; Kulkarni 2013), started in 2013. The next-generation Palomar time-domain survey, the Zwicky Transient Facility (ZTF; Kulkarni 2012), will begin in 2017. ZTF will upgrade the camera on P48 and feature improved electronics and a ~ 47 deg² field of view (FOV), which is a factor of ~ 7 increase over the PTF FOV.

A primary motivation for both PTF and ZTF is the search for fast ($\lesssim 24$ hr) transients, a rare class of explosive events expected to include “kilonovae,” the result of binary neutron star (BNS) mergers (e.g., Kasen et al. 2015). BNS mergers are thought to be the most promising electromagnetic counterparts to gravitational wave (GW) sources (e.g., Metzger & Berger 2012; Nissanke et al. 2013). Now that we are firmly in the age of GW detections (Abbott et al. 2016b), the search for electromagnetic counterparts is both highly exciting and extremely pressing. As surveys identify fast-transient candidates, including GW counterparts, they contend with significant foreground contamination in the form of stellar flares (e.g., Kulkarni & Rau 2006; Berger et al. 2012). The systematic removal of faint stars from extragalactic candidate lists can fully alleviate this problem by removing false positives (FPs) from consideration for expensive follow-up resources. Indeed, while searching for an optical counterpart to GW150914, a (now outdated) PTF star catalog rejected $\sim 40\%$ of the viable transient candidates (Kasliwal et al. 2016).

PTF employs sophisticated software solutions to rapidly process new observations, perform image subtraction, and identify transient candidates (Cao et al. 2016; Masci

⁶ Hubble Fellow.

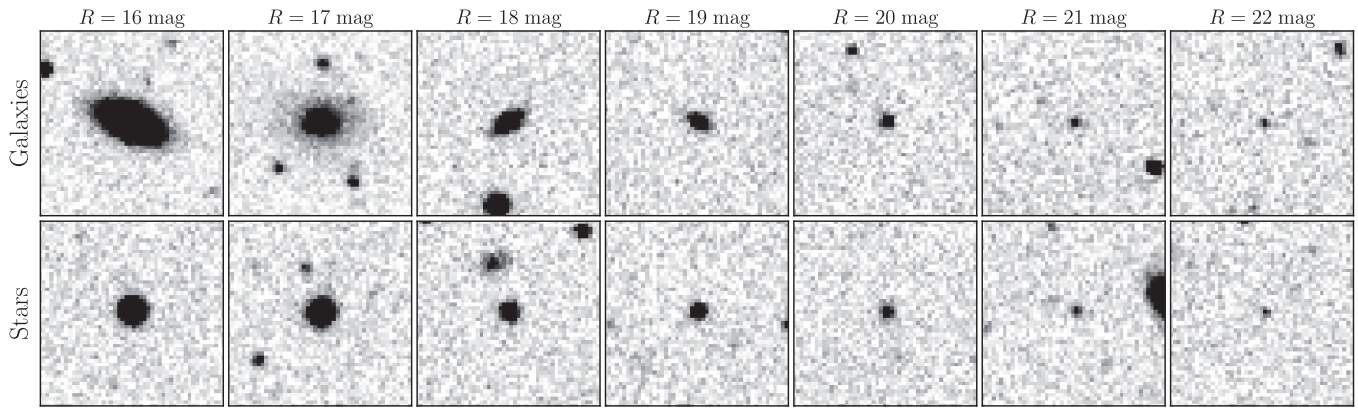


Figure 1. Postage stamps showing typical stars and galaxies in PTF reference images as a function of magnitude. The images show that stars and galaxies can easily be separated by eye down to $R \approx 19$ mag, while for fainter sources the two are virtually indistinguishable. Postage stamps are 50×50 pixels, centered on the source of interest, with north up and east to the left. Source classifications are from SDSS spectra. Each stamp is from a reference co-add of five individual PTF images, the shallowest reference images produced by PTF, yielding an effective exposure time of 300 s.

et al. 2016). These candidates are then confirmed or rejected as bona fide astrophysical variations by machine-learning software (e.g., Bloom et al. 2012; Brink et al. 2013; Rebbapragada et al. 2015). At this stage human vetting of the candidates identifies those that merit additional follow-up observations. Within the Sloan Digital Sky Survey (SDSS; York et al. 2000) imaging footprint stars and galaxies can be separated with high fidelity to a faintness of ~ 22 mag, by comparing the point-spread function (PSF) magnitude with the best-fit model magnitude.⁷ However, SDSS only overlaps approximately half of the full PTF imaging footprint, and faint objects cannot reliably be classified as stars or galaxies via visual inspection in PTF images, as illustrated in Figure 1.

It naturally follows that the development of a star–galaxy separation model for PTF would improve our ability to reject FPs in our search for fast transients and GW counterparts. The optimal star–galaxy catalog for fast-transient surveys would identify as many stars as possible (true positives [TPs]), while minimizing the number of galaxies misclassified as stars (FPs). Striking the proper balance between these two objectives is challenging: an overly conservative selection of stars will result in many transient candidates with Galactic origin, while an overly aggressive selection will lead to many galaxies being excluded from the search. The intrinsic rarity of fast transients and GW counterparts means that the latter situation, which could result in a GW counterpart being missed entirely, is especially undesirable. Machine-learning algorithms offer an attractive solution to this problem as they enable a precise tuning of the classification decision threshold to balance the number of TPs and FPs.

Supervised machine-learning algorithms construct a model to map *features*, measured properties of the sources, to *labels*, such as a classification or physical property.⁸ The model is constructed using a training set, and its performance is evaluated using a test set. The training set and test set are independent subsets of the data with (spectroscopic) labels that we adopt as ground truth. Machine-learning models are very flexible, capable of capturing complex nonlinear behavior in the multidimensional feature space. In many cases they provide

fast, automated classifications for new data. Previously, machine-learning models utilizing decision trees have been used to successfully classify stars and galaxies in SDSS imaging data (Ball et al. 2006; Vasconcellos et al. 2011).

Here, we construct a model employing an ensemble of decision trees, trained with spectroscopic classifications from SDSS, to separate stars and galaxies in PTF images. We describe our procedure to curate an appropriate training set and the steps utilized to optimize the performance of the algorithm. Most importantly, we compare the performance of our model to that of *SExtractor*, which currently provides the best discriminant between stars and galaxies in PTF images outside the SDSS photometric footprint. We define conservative selection criteria for stellar classification and apply the final optimized model to all $>5 \times 10^8$ sources in the PTF photometric catalog. This catalog has been ingested by the appropriate PTF pipelines and is currently used to reject FPs in the search for new transients.

2. TRAINING THE MODEL WITH SDSS SPECTROSCOPIC TARGETS

An important and essential first step in the construction of a supervised machine-learning model is the curation of the training set and test set. The data-driven nature of supervised machine learning means that special consideration must be taken to avoid potential biases in the training set. The final model predictions for the full data set will reflect, and likely preserve, any biases in the training set. For the PTF star–galaxy catalog, features are extracted from PTF reference images (deep co-adds) using *SExtractor* (Bertin & Arnouts 1996), and labels are provided by SDSS spectroscopic observations.

2.1. SDSS Training Labels

To facilitate the search for transient sources, the PTF imaging pipeline produces reference images (Laher et al. 2014), deep co-adds of ≥ 5 individual 60 s exposures. PTF reference images are significantly deeper and offer superior image quality to individual exposures. We employ only the *R*-band detections for the model because there are significant gaps in the sky coverage for the other PTF filters. To train our model, we consider all photometric detections from PTF *R*-band reference images. Using all *R*-band references available as of 2016 July 22 UT, there are 548,687,903 sources detected by *SExtractor*.

⁷ See http://www.sdss.org/dr12/algorithms/classify/#photo_class for further details.

⁸ For a more detailed primer on machine learning, we refer the reader to Hastie et al. (2009).

PTF employs a grid of overlapping pointings; thus, some of those ~ 550 million detections represent duplicates of the same astrophysical source.

To identify which photometric detections are suitable to train the machine-learning model, we adopt the labels from SDSS spectroscopic classifications as “ground truth.” Optical spectra taken as part of the original SDSS and the Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013) were automatically classified as belonging to one of three classes: stars, galaxies, and quasi-stellar objects (QSOs). Using PTF imaging data, we hope to separate resolved (galaxies) and unresolved (stars, QSOs) sources, which for simplicity will be hereafter referred to as galaxies and stars, respectively.

Using the spatial cross-match tool available via SDSS *CasJobs*, all PTF photometric sources coincident within $1''$ of an SDSS spectroscopic source are selected as potential training objects. In total, there are 3,193,349 matches between PTF and the SDSS spectroscopic catalog. To prevent overfitting, these sources were split roughly 60–40 into independent training and test sets. The training and test sets are used to optimize the model and evaluate its accuracy, defined as the fraction of sources that are correctly classified, respectively. As previously mentioned, there are photometric duplicates in the PTF reference-image catalogs. In addition to this, SDSS obtained spectra of some sources more than once. Thus, a random 60–40 split of the ~ 3 million training objects would not ensure independence between the training and test sets. To prevent sources from being assigned to both sets, we randomly select 60% of the unique *objid*, the SDSS photometric identification key, and assign all sources with matching *objid* to the training set. All remaining sources are assigned to the test set. Following this procedure, the training set includes 1,919,088 sources, while the test set has 1,274,261 sources.

Qualitatively, the distribution of the number of co-adds, N_{coadd} , in the reference image on which a source is detected is similar for the full ~ 550 million PTF source catalog and the ~ 3 million training sources. For both the full catalog and the training set a plurality of sources have $N_{\text{coadd}} = 5$, 46% and 38%, respectively. Both distributions exhibit strong positive skew, with a secondary peak at $N_{\text{coadd}} = 50$, the maximum number of co-adds. Ultimately, the training set is more biased toward deep images, with 11% of sources having $N_{\text{coadd}} \geq 30$, while the same is true for only 5% of sources in the full catalog. Nevertheless, we do not expect these differences to produce significant biases in the final star–galaxy predictions because the overall distributions are similar, and the training set is slightly deeper and less noisy.

2.2. SExtractor Photometric Features

The PTF reference-image pipeline utilizes SExtractor to measure 96 photometric properties per source. Relevant features for classifying sources as either stars or galaxies include elongation, FWHM, best-fit Petrosian radius, etc. Several properties measured by SExtractor are contextual, such as the X and Y positions of the source photocenter on the CCD, and we exclude these from the machine-learning model. Furthermore, we normalize all SExtractor shape measurements by the average seeing in a given image⁹ and all flux

measurements by the flux in a circular aperture with 2 pixel diameter. The former accounts for the variable observing conditions for different references, while the latter helps to remove biases due to the brightness distribution of SDSS spectroscopic targets (see Section 4).¹⁰

A list of SExtractor measurements that are excluded from the machine-learning model is given in Table 1.¹¹ The majority of the features in this table are uninformative. One major exception is *CLASS_STAR*, which is a neural-network-based source classification ranging from 0 to 1. Sources with *CLASS_STAR* ≈ 1 are considered star-like, while sources with *CLASS_STAR* ≈ 0 are non-star-like (galaxies, but also cosmic rays, etc.). Outside the SDSS photometric footprint, *CLASS_STAR* represents the best model for separating stars and galaxies in PTF data. Thus, we exclude *CLASS_STAR* from the model so that we may compare our final classifications against those made by SExtractor.

Features provided to the machine-learning model are listed in Table 2. As previously noted, shape parameters are normalized by the average seeing in a reference image, while all magnitude measurements are normalized relative to the magnitude measured in a 2 pixel diameter circular aperture. We exclude the uncertainties on the shape and brightness measurements from the model as these primarily reflect the depth of the reference image, which varies considerably over the data set given that some co-adds include five images while others include 50. Following normalization, we supply the machine-learning model with 43 features. A kernel density estimate (KDE)¹² of the probability distribution function (PDF) of five uncorrelated features is shown in Figure 2.

2.3. Removal of Photometric Blends

During the final stages of model construction, we noticed an unusual systematic whereby a large fraction of stars with $r' \approx 20.5$ mag were erroneously classified as galaxies (see Section 4 and Figure 5 for further details). Manual inspection of several of these sources revealed that they were red stars blended with fainter sources. The SDSS spectroscopic survey was intentionally biased toward observing luminous red galaxies (LRGs; see, e.g., Eisenstein et al. 2001), and thus faint red stars that are photometrically blended would satisfy the general LRG selection criteria of being red, faint, and extended.

We found that these sources could be readily identified using both the *class* and *sourceType* columns in the *specObjAll* table of the DR12 SDSS database (Alam et al. 2015). These columns elucidate the spectroscopic class (GALAXY, QSO, or STAR) of the target and the reason the source was targeted, respectively. Our ultimate goal with this classification catalog is to develop a pristine list of point sources. In other words, we are willing to accept stars being classified as galaxies if those stars are blended with other sources such that their photometric appearance resembles galaxies. While such an approach would be disadvantageous

⁹ For PTF the seeing is determined from a trimmed mean of the *FWHM_IMAGE* parameter measured by SExtractor (Laher et al. 2014).

¹⁰ While contextual information, such as brightness or galactic latitude, could in principle help data-driven classification, in many cases contextual features propagate biases in target selection to the final model (see, e.g., Richards et al. 2012). Hence, we exclude positional coordinates and normalize brightness measurements for our final model.

¹¹ For a full description of all SExtractor features see the [documentation](#) or Dr. Benne Holwerda’s excellent [Guide to SExtractor](#).

¹² All KDEs presented in this paper adopt a Gaussian kernel and Scott’s rule to determine the kernel bandwidth (Scott 1992).

Table 1
PTF SExtractor Features Excluded from the Model

Name	Description
NUMBER	Identification number of object
X_IMAGE, Y_IMAGE	Pixel position of object centroid
XWIN_IMAGE, YWIN_IMAGE	Pixel position of object centroid, wind-owed measurement
X_WORLD, Y_WORLD	R.A. and decl. coordinates of object centroid
XPEAK_IMAGE, YPEAK_IMAGE	Pixel position with peak object intensity
ALPHAWIN_J2000	Right ascension of object barycenter (J2000)
DELTAWIN_J2000	Declination of object barycenter (J2000)
THETAWIN_J2000	Object position angle (east of north) (J2000)
CLASS_STAR	SExtractor stellarity index between 0 and 1; 1 = star
MAG_ISO	Isophotal magnitude measurement
MAGERR_ISO	
MAG_ISOCOR	Corrected isophotal magnitude
MAGERR_ISOCOR	
MAG_BEST	Best of MAG_AUTO and MAG_ISOCOR
MAGERR_BEST	
MU_THRESHOLD	Surface brightness detection threshold above background
BACKGROUND	Background at object centroid position
THRESHOLD	Detection threshold above background
ISOAREA_WORLD	Isophotal area above threshold
ISOAREAF_WORLD	Isophotal area (filtered) above threshold (degrees)
ISO0, ISO1, ISO2, ISO3, ISO4, ISO5, ISO6, ISO7	Isophotal area at level n
FLUX_BEST	Best of FLUX_AUTO and FLUX_ISOCOR
FLUXERR_BEST	
FLUX_AUTO	Flux within a Kron-like elliptical aperture
FLUXERR_AUTO	
FLUX_ISO	Isophotal flux
FLUXERR_ISO	

Note. Feature names including “ERR” indicate uncertainty measurements for the immediately preceding feature. The use of isophotal magnitude measurements has been deprecated in SExtractor; therefore, we exclude these features (MAG_ISO, MAG_ISOCOR, MAG_BEST) from the final model.

to galaxy clustering studies, it is ideal for the search for transients. Thus, we exclude spectroscopic stars targeted as galaxies from the training set. Similarly, we exclude spectroscopic galaxies targeted as either stars or QSOs and spectroscopic QSOs targeted as galaxies. We additionally exclude all spectroscopic QSOs with redshift $z < 1$, many of which have detectable host galaxies in addition to their active galactic nuclei. Finally, we exclude a small number (34,437) of emission-line galaxy (ELG) and Sloan Extended Quasar, ELG, and LRG Survey (SEQUELS) targets, which consistently have spectroscopic classes that do not match their target class.

The full list of SDSS spectroscopic class and target type combinations that were excluded from the training and test sets is summarized in Table 3. In total, these exclusions remove 194,509 sources from the 3,193,349 PTF sources with SDSS spectra. The final training and test sets, which we hereafter refer to as *photometrically clean*, include 1,802,357 and 1,196,483 sources, respectively.

3. MACHINE-LEARNING MODEL CONSTRUCTION

3.1. The Random Forest Algorithm

Random forest (RF) methods utilize the aggregation of multiple decision trees to assign a final classification or regression value to newly observed sources (Breiman 2001). RF makes use of bagging (see Breiman 1996), wherein bootstrap samples of the training set are used to construct each of the N_{tree} total trees in the forest. As each tree in the forest is constructed, only a random subset of m_{try} features is selected from the full feature set as a potential splitting criterion at each node of the tree. The use of bagging and m_{try} random features reduces the variance of the final model relative to single decision-tree models, providing low-bias, low-variance predictions. The final RF predictions are determined by averaging the predictions for a new source from each of the N_{tree} individual trees. Furthermore, the RF algorithm is fast, each of the trees can be constructed independently and thus in parallel, and they are relatively easy to interpret. RF models have recently become highly popular as an application for astronomical data sets, due to their relative insensitivity to noisy or meaningless features (e.g., Brink et al. 2013; Miller et al. 2015) and their invariant response to even highly non-Gaussian feature distributions (e.g., Dubath et al. 2011; Richards et al. 2011). Due to its flexibility and speed, we adopt RF for this study, in particular, we utilize the Python scikit-learn¹³ implementation of the algorithm (Pedregosa et al. 2011).

3.2. Imputation for Missing Features

An initial challenge for the classification model is that SExtractor does not always produce finite measurements for the features listed in Section 2.2. For a small number of sources BWIN_WORLD is reported as NaN, while a slightly larger number of sources have APER_FLUX measurements of 0, which results in a normalized feature value of infinity (see Table 2). In the training set, a single source has a bad BWIN_WORLD measurement, while 48, 47, 47, 46, 45, and 44 sources have zero-valued aperture flux measurements from the smallest to the largest aperture, respectively. In the full, 548,687,903-source PTF reference catalog 731 sources have bad BWIN_WORLD measurements, while 11,958, 11,421, 11,160, 10,687, 10,595, and 10,474 sources have zero-valued aperture flux measurements from the smallest to the largest aperture, respectively.

There are three potential solutions to deal with missing features: (i) exclude any features with missing data from the model entirely, (ii) exclude the sources with missing features from the model training and final predictions, or (iii) develop a method to estimate the values of the missing features. The first two options are undesirable as they remove information from the model and prevent predictions for some sources, respectively. The third option is most attractive as it does not exclude any valuable information.

We test two methods of imputation to estimate the missing values in the feature set. The first is simple: replace all missing values with the median value of the feature in the training set. The second is more complex: use RF regression to perform a nonparametric estimate of the missing features using the features with no missing values. In particular, we use RF regression with $N_{\text{tree}} = 100$ and fully grown trees to estimate the missing values. Stekhoven & Bühlmann (2012) have shown

¹³ <http://scikit-learn.org/stable/>

Table 2
PTF SExtractor Features Included in the Model

Name	Description	Normalization Factor
X2_IMAGE	Second-order moment of object, along x -axis	$1/\text{seeing}^2$
Y2_IMAGE	Second-order moment of object, along y -axis	$1/\text{seeing}^2$
X2WIN_IMAGE	Second-order moment of object, windowed measurement	$1/\text{seeing}^2$
ERRX2WIN_IMAGE		
Y2WIN_IMAGE	Second-order moment of object, windowed measurement	$1/\text{seeing}^2$
ERRY2WIN_IMAGE		
XY_IMAGE	Covariance of position between x and y	$1/\text{seeing}^2$
XYWIN_IMAGE	Covariance of position between x and y , windowed measurement	$1/\text{seeing}^2$
ERRXYWIN_IMAGE		
AWIN_WORLD	Object profile rms along the major axis, windowed measurement	$1/\text{seeing}$
ERRAWIN_IMAGE		
BWIN_WORLD	Object profile rms along the minor axis, windowed measurement	$1/\text{seeing}$
ERRBWIN_IMAGE		
MAG_APER ₂	Magnitude in a 2 pixel diameter circular aperture centered on the object	See table notes
MAGERR_APER ₂		
MAG_APER ₄	Magnitude in a 4 pixel diameter circular aperture centered on the object	MAG_APER ₂
MAGERR_APER ₄		
MAG_APER ₅	Magnitude in a 5 pixel diameter circular aperture centered on the object	MAG_APER ₂
MAGERR_APER ₅		
MAG_APER ₈	Magnitude in an 8 pixel diameter circular aperture centered on the object	MAG_APER ₂
MAGERR_APER ₈		
MAG_APER ₁₀	Magnitude in a 10 pixel diameter circular aperture centered on the object	MAG_APER ₂
MAGERR_APER ₁₀		
MAG_APER ₁₄	Magnitude in a 14 pixel diameter circular aperture centered on the object	MAG_APER ₂
MAGERR_APER ₁₄		
MAG_AUTO	Kron-like elliptical aperture magnitude	MAG_APER ₂
MAGERR_AUTO		
MAG_PETRO	Petrosian-like elliptical aperture magnitude	MAG_APER ₂
MAGERR_PETRO		
MU_MAX	Peak surface brightness above background	MAG_APER ₂
THETA_IMAGE	Position angle of object, counterclockwise	
THETAWIN_IMAGE	Position angle of object, counterclockwise, windowed measurement	
ERRTHETAWIN_IMAGE		
THETAWIN_WORLD	Position angle of object, counterclockwise, world coordinates	
ELONGATION	A_IMAGE/B_IMAGE	
FWHM_IMAGE	FWHM of object, assuming Gaussian core	$1/\text{seeing}$
KRON_RADIUS	Kron radius of object	$1/\text{seeing}$
PETRO_RADIUS	Petrosian radius of object	$1/\text{seeing}$
ISOAREAF_IMAGE	Isophotal area (filtered) above threshold	$1/\text{seeing}^2$
FLUX_APER ₂	Flux in a 2 pixel diameter circular aperture centered on the object	$1/\text{FLUX_MAX}$
FLUXERR_APER ₂		
FLUX_APER ₄	Flux in a 4 pixel diameter circular aperture centered on the object	$1/\text{FLUX_MAX}$
FLUXERR_APER ₄		
FLUX_APER ₅	Flux in a 5 pixel diameter circular aperture centered on the object	$1/\text{FLUX_MAX}$
FLUXERR_APER ₅		
FLUX_APER ₈	Flux in an 8 pixel diameter circular aperture centered on the object	$1/\text{FLUX_MAX}$
FLUXERR_APER ₈		
FLUX_APER ₁₀	Flux in a 10 pixel diameter circular aperture centered on the object	$1/\text{FLUX_MAX}$
FLUXERR_APER ₁₀		
FLUX_APER ₁₄	Flux in a 14 pixel diameter circular aperture centered on the object	$1/\text{FLUX_MAX}$
FLUXERR_APER ₁₄		
FLUX_RADIUS ₂₅	Radius enclosing 25% of the object flux	$1/\text{seeing}$
FLUX_RADIUS ₅₀	Radius enclosing 50% of the object flux	$1/\text{seeing}$
FLUX_RADIUS ₈₅	Radius enclosing 85% of the object flux	$1/\text{seeing}$
FLUX_RADIUS ₉₅	Radius enclosing 95% of the object flux	$1/\text{seeing}$
FLUX_RADIUS ₉₉	Radius enclosing 99% of the object flux	$1/\text{seeing}$
FLUX_MAX	Flux of brightest pixel within the object	See table notes
FLAGS	Source Extractor Flags, coded in bitmask	

Note. Feature names including “ERR” indicate uncertainty measurements for the immediately preceding feature. These measurements of the uncertainty are not used by the model. Normalization factors are multiplicative, aside from mag measurements, and are applied, as needed, prior to running the model. Magnitudes are logarithmic; thus, all mag measurements are normalized via a difference (e.g., the Petrosian mag is represented as $\text{MAG_APER}_2 - \text{MAG_PETRO}$). In practice, the inverse of the aperture flux is used (e.g., the flux in a 5 pixel aperture is represented as $\text{FLUX_MAX}/\text{FLUX_APER}_5$). Each of the eight SExtractor FLAGS is treated as a binary feature for the classifier. MAG_APER_2 and FLUX_MAX are not represented in the final model as they are otherwise captured as normalization factors. In sum, 43 features are included in the machine-learning model.

that this nonparametric method outperforms several other common methods for imputation.

To test which of these two methods works best for the features with missing values, we perform a threefold cross-validation

(CV) run on the training set to estimate the value of the seven aforementioned features for every source where the feature value is not missing. Thus, we can compare the imputation estimate with the true value and evaluate which of the two methods is

superior via the rms error (RMSE). The results of this test are shown in Table 4. The performance of the two methods is similar for each of the aperture flux features; however, the RF regression method clearly provides superior predictions for BWIN_WORLD. We will later show that BWIN_WORLD is important for star-galaxy classification; thus, we adopt the RF regression method for feature imputation.

3.3. Feature Selection

In addition to the curation of the training set and choice of algorithm, feature engineering is an important step during machine-learning model construction. As previously noted, RF methods are relatively insensitive to weak or uninformative features, and they also perform well in the presence of strongly correlated features (e.g., Richards et al. 2012). Nevertheless, we test the full feature set to see whether the model performance can be improved by removing some features.

RF methods naturally provide a measure of relative feature importance: features that are preferentially selected near the top of individual decision trees contribute more to the final classification predictions than features near the bottom of the trees. Aggregating this information over all trees provides a measure of relative importance for the individual features. In the presence of highly correlated features, this method does not provide perfectly accurate results as the correlated features may replace each other at the top of the tree, thereby suppressing their relative importance. Nevertheless, we employ the RF feature importance rankings to determine which features, if any, can be removed from the model.

Our procedure is similar to the one employed in Richards et al. (2012). We construct a series of RF models whereby we iteratively add one feature at a time to each successive model starting with the most-important-RF feature and ending with the 41st most important.¹⁴ We assess the accuracy of each model via a threefold CV run on the training set, and we repeat this procedure five times to estimate the scatter in the performance of each model. The results of this procedure are shown in Figure 3, which shows that only three features are needed to achieve a classification accuracy within $\sim 1\%$ of the maximum CV accuracy. The gains in accuracy beyond the first three features are marginal, but increasing, through the 37th feature, FLAG1, after which the accuracy decreases slightly. Thus, we select all the ranked features up to FLAG1 for use by the final classification model. The exclusion of the last four features does not significantly alter the final model predictions.

Finally, we note that we explored the procedure described in Dubath et al. (2011) to only include uncorrelated features in the final RF model. This procedure produced a maximum CV accuracy of 0.965, whereas the method described above produces a maximum CV accuracy of 0.980. Given the $\sim 1.5\%$ improvement in prediction accuracy, we elect to include the correlated features in the final model.

3.4. Optimizing the Model Tuning Parameters

The RF algorithm has multiple tuning parameters, which, in combination, control the smoothness of the model projection in the multidimensional feature space. The two most important

tuning parameters, N_{tree} and m_{try} , were previously mentioned in Section 3.1. Additional tuning parameters control the depth of individual trees in the forest. We optimize the `nodesize` parameter, which prevents further splitting of the tree if it would result in a terminal node with fewer than `nodesize` sources.

To optimize the model, we perform a coarse-grid search over the three tuning parameters. At each point on the grid, we perform threefold CV on the training set to evaluate the model accuracy for the given tuning parameters. We further refine the tuning parameters using a fine-grid search centered on the optimal model from the coarse-grid search. The final optimized model has a CV accuracy of 0.98, with parameters $N_{\text{tree}} = 750$, $m_{\text{try}} = 14$, and `nodesize` = 1. We note that the final model predictions are not sensitive to the tuning parameters: the worst model from the fine-grid search is $< 0.1\%$ worse than the optimal model.

4. EVALUATION OF THE OPTIMIZED MODEL

4.1. PTF RF, CLASS_STAR, and SDSS Comparison

To test the accuracy of the final, optimized model, we train an RF using the training set and the optimized tuning parameters from Section 3.4. This model is then applied to the independent test set, where we can compare the model predictions with spectroscopic classifications. For the test set, the RF model produces an overall prediction accuracy of 98.0%, which represents a $\sim 3.9\%$ improvement over the accuracy of the SExtractor stellerity measure CLASS_STAR (94.4%).¹⁵ More impressive, however, is the performance of the RF model on faint sources. For test set sources with $r' \geq 21$ mag, CLASS_STAR has an accuracy of 77.2%, while the RF model has an accuracy of 92.0%. This represents an improvement of $\sim 19\%$ for the faintest sources detected by PTF.

Interestingly, neither method performs as well as the simple parametric method employed by the SDSS pipeline. In brief, the SDSS pipeline identifies all sources with $\text{psfMag} - \text{cModelMag} > 0.145$ as galaxies (Lupton et al. 2002), where `psfMag` is the PSF magnitude and `cModelMag` is the composite-model magnitude resulting from the best-fitting linear combination of the best-fitting de Vaucouleurs and exponential model for an object's light profile. For the test set, the SDSS photometric classification provides an overall accuracy of 99.6% and an accuracy of 99.1% for sources with $r' \geq 21$ mag. We attribute the improved performance of the simplistic SDSS photometric classifier to their higher-quality observations, including better seeing of $\sim 1''.4$ for SDSS (Abazajian et al. 2003) versus $\sim 2''.4$ for PTF and greater depth for SDSS.¹⁶

In Figure 4, we compare the accuracy of the RF model, CLASS_STAR, and the SDSS photometric classifier evaluated via the test set as a function of magnitude. Figure 4 shows that the performance of each method is similar down to $r' \approx 18$ mag. This is to be expected based on Figure 1, which shows that galaxies and stars are clearly separable over this magnitude

¹⁴ Three of the SExtractor flags were identified as having zero importance by RF, so we excluded them from this exercise. We further added a uninformative feature, `NoInfo`, which is identically 0 for all sources. The inclusion of `NoInfo` can help to identify noisy features (see Brink et al. 2013).

¹⁵ Overall model accuracies are evaluated using a threshold of 0.5 for separating stars and galaxies. Thus, an RF probability > 0.5 or CLASS_STAR > 0.5 results in a stellar classification for the PTF RF model and SExtractor, respectively.

¹⁶ The SDSS photometric classifier uses the sum of `psfMag` - `cModelMag` across all five filters to perform the final star-galaxy classification.

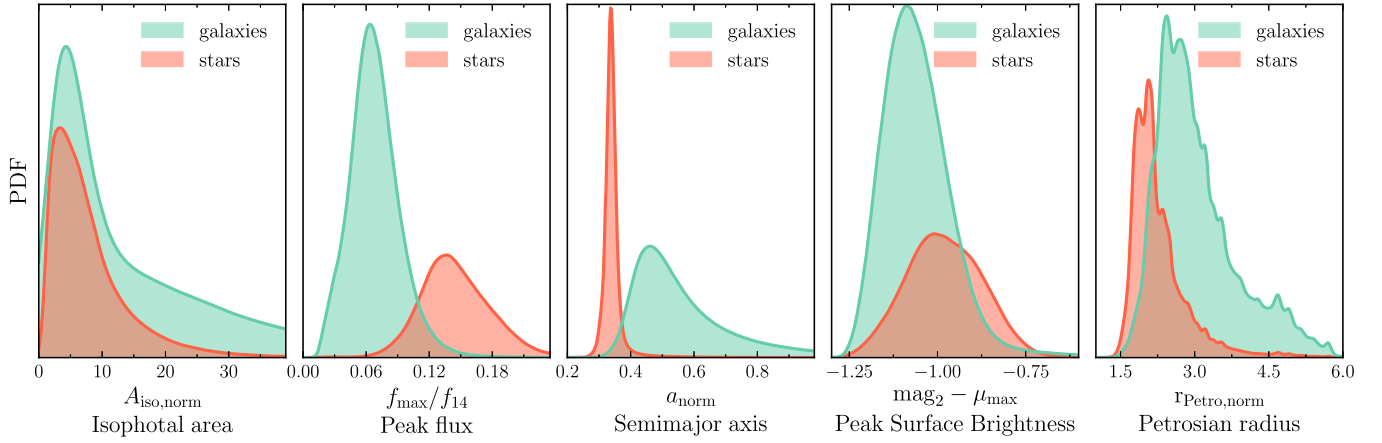


Figure 2. KDE of the PDF for select model features for all PTF sources with spectroscopic observations. The area under the curves has been scaled relative to the total number of stars and galaxies in the spectroscopic set (galaxies outnumber stars by a factor of ~ 2.1). Distributions are for the normalized features (see the text and Table 2 for further details). While no single feature separates stars and galaxies significantly better than `CLASS_STAR`, the differing PDFs for the two populations suggest that a nonparametric method may produce a significant improvement over the `SExtractor` stellarity estimate.

range in PTF imaging. The performance of `CLASS_STAR` quickly degrades for fainter sources, however, to the level that `CLASS_STAR` is similar to random guessing for sources with $r' > 22$ mag. As previously noted, the RF model provides superior predictions for faint sources, which will enable us to better identify stars in PTF imaging outside the SDSS footprint. Figure 4 also shows a KDE of the PDF of $N(m)$ for training set sources, which is virtually identical to the PDF for test set sources. The RF model is most reliable in regions of high density, roughly $16.5 \text{ mag} \lesssim r' \lesssim 21.3 \text{ mag}$.

4.2. RF Model Accuracy for Stars and Galaxies

As previously noted, the primary motivation for constructing a PTF star–galaxy catalog is to identify a pristine list of point sources in PTF imaging. We are particularly interested in the accuracy with which we can identify faint stars as these are the most likely FPs in the search for fast transients (Kulkarni & Rau 2006; Berger et al. 2013). Similar to above, we plot the accuracy of the RF model for classifying stars and galaxies in the test set as a function of brightness in Figure 5.

Figure 5 features four panels, each of which reflects slight variations on the training set. The left column shows training sets that include all PTF sources with SDSS spectra, while the right column shows results from the *photometrically clean* training set (see Section 2.3). The top row shows training sets including all available stars and galaxies, while the bottom row shows the results when the stars and galaxies in the training set are downsampled such that both classes have similar PDFs of magnitude.

The upper left panel contains the most striking feature in Figure 5: the kink in the accuracy curve for stars at $r' \approx 20.7$ mag, followed by the crossing of the star and galaxy accuracy curves at $r' \approx 21.5$ mag. A less significant, but nonetheless noticeable, kink also appears near $r' \approx 17.5$ mag in the star accuracy curve. These departures from a smooth accuracy curve occur near peaks in the Galaxy PDF, where stars are most significantly outnumbered.

Initially, we believed that the kinks could be removed by balancing the magnitude PDF for stars and galaxies in the training set. To achieve this balance, we use KDEs of the PDFs from $r' = 13.5$ mag to $r' = 22.5$ mag. The stellar KDE and galaxy KDE are evaluated at the brightness of each galaxy, and

the former is divided by the latter to provide a weight. We then select a weighted random sample of 500,000 galaxies for the balanced training set. We use the same procedure to select a weighted random sample of 500,000 stars; however, the weights are determined by dividing the Galaxy PDF by the stellar PDF. All training set sources with brightness outside the range $r' = 13.5\text{--}22.5$ mag are also included, resulting in a final balanced training set with 1,001,975 sources. The balanced training set PDFs and accuracy curves are shown in the lower left panel of Figure 5. While the significance of the kinks is reduced when using the balanced training set, it is clear that balancing the two classes does not eradicate this unusual systematic behavior.

Ultimately, the kink in the upper left panel is due to the SDSS targeting bias toward LRGs, a small fraction of which turn out to be photometrically blended red stars, as previously noted in Section 2.3. In brief, the removal of stars targeted as galaxies, galaxies targeted as stars, and low- z QSOs from the training and test sets dramatically improves the performance of the star–galaxy classification model.¹⁷ This also removes the unusual systematics from the accuracy curves, as seen in the right column of Figure 5. We applied the same procedure described above to balance the *photometrically clean* training set, and the resulting predictions are shown in the lower right panel of Figure 5. Ultimately, the performances of the full and balanced *photometrically clean* training sets were nearly identical, with an overall accuracy of 98.0% and 97.8%, respectively. Ultimately, we adopt the full *photometrically clean* training set for the final RF model as this provides the most information to the classifier. The use of the balanced *photometrically clean* training set would not significantly alter the final model classifications.

4.3. Selecting a Pristine Sample of Stars

Rather than producing the best overall accuracy possible, we hope to generate a catalog of PTF point sources that is virtually free of galaxies. In addition to providing classifications, RF models also produce relative rankings of the class likelihood for newly observed sources by recording the fraction of trees in

¹⁷ For the full details on which sources are removed from the training and test sets, see Section 2.3.

Table 3
SDSS Spectroscopic Targets Excluded from the Training Set

Class	STAR	QSO	GALAXY
sourceType	LRG (16444)	SEQUELS_TARGET (16,511)	QSO (29,106)
	GALAXY (11142)	LRG (16,150)	SEQUELS_TARGET (12,210)
	HIZ_LRG (169)	GALAXY (3293)	SEQUELS_ELG (1545)
		SN_GAL1 (1297)	FAINT_ELG (1071)
		SEQUELS_ELG (528)	SEQUELS_ELG_LOWP (986)
		ELG (525)	ELG (887)
		SEQUELS_ELG_LOWP (174)	STAR (812)
			WISE_BOSS_QSO (694)
			QSO_EBOSS_W3_ADM (548)
			QSO_WISE_FULL_SKY (371)
			QSO_VAR_SDSS (362)
			QSO_VAR_LF (324)
			QSO_WISE_SUPP (269)
			SERENDIPITY_BLUE (262)
			SERENDIPITY_DISTANT (237)
			QSO_GRI (227)
			QSO_DEEP (147)
			STD (128)
Total	27,755	116,568 ^a	50,186

Note. For each spectroscopic class (Class) the corresponding number of sources from each sourceType that are removed from the training and validation set combined is shown in parentheses.

^a In addition to the 38,478 spectroscopic QSOs detailed above, an additional 78,090 spectroscopic QSOs with $z < 1$ are removed from the training and validation sets.

which each source is assigned to each class.¹⁸ For our two-class problem, a source that is classified as a star in every tree would have an RF relative ranking equal to 1, while a source labeled a galaxy in every tree would have ranking 0. Sources with RF relative ranking ≈ 0.5 are somewhat ambiguous, with the trees nearly divided on the classification.

Thresholds can be placed on the RF relative ranking to adjust the overall number of FPs and TPs produced by the classifier. Above, a threshold of 0.5 was adopted to test the overall accuracy of the classifier. Now, we adjust that threshold to reduce the number of FPs, which for our purposes are considered far more harmful than false negatives (FNs). Threshold adjustments are typically determined using a receiver operating characteristic (ROC) curve, which plots the true positive rate (TPR),

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

against the false positive rate (FPR),

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}},$$

where TN is the number of true negatives, as the classification threshold is varied from 1 to 0. The performance of different models can be compared via ROC curves by examining which comes the closest to the classification ideal of $\text{TPR} = 1$ and $\text{FPR} = 0$.

ROC curves for the PTF RF model and SExtractor are shown in Figure 6. For SExtractor the curve is determined by varying the classification threshold from $\text{CLASS_STAR} = 0$

to 1. The performance of SDSS is shown as a single point because the SDSS pipeline provides a single binary classification without any information on the relative likelihood for individual sources. Similar to Figure 4, Figure 6 shows that the PTF RF model significantly outperforms the SExtractor model. In fact, the performance of the PTF RF model on faint sources ($r' > 20.5$ mag) is virtually identical to the performance of SExtractor on the entire test set. As has already been noted, the superior quality of SDSS imaging results in higher-fidelity classifications than is possible with PTF imaging.

The classification threshold adopted for the PTF point-source catalog is determined by maximizing the TPR at $\text{FPR} = 0.005$. The adoption of this low FPR ensures that less than 0.5% of galaxies will be included in the point-source catalog and thereby excluded from examination should they host a transient. For the test set at $\text{FPR} = 0.005$, the PTF RF model produces $\text{TPR} = 0.91$, corresponding to a classification threshold of 0.83 for the RF relative ranking. Below, we show that the performance of the model as measured by the test set likely overstates the model accuracy when applied to sources in the field. Thus, we ultimately adopt a classification threshold that is more conservative than 0.83.

5. IMPLEMENTING THE CATALOG

5.1. Final Field-source Predictions

The final step for incorporating the star–galaxy catalog into the appropriate PTF pipelines is to apply the RF model to the 548,687,903 sources detected in PTF reference images. To assess the efficacy of the model as applied to the field star population, which is dominated by sources at the faint end of the test set, we compare our final PTF classifications to those made by the SDSS photometric pipeline, which for this purpose we adopt as ground truth.

¹⁸ These relative rankings are often referred to as RF probabilities. However, the RF probability score does not represent a true probability as it is a strong function of the training set, which in virtually all astronomical applications is biased relative to the true distributions present in nature. Thus, we prefer to refer to this quantity as the RF relative ranking.

Table 4
Imputation Results

Method	BWIN_WORLD	FLUX_APER ₂	FLUX_APER ₄	FLUX_APER ₅	FLUX_APER ₈	FLUX_APER ₁₀	FLUX_APER ₁₄
RF	0.028	6411.8	1589.8	1012.4	13.2	8.6	4.7
Median	0.183	6400.2	1599.7	1017.3	14.3	8.9	4.9

Note. Table columns show the RMSE when comparing the imputation predictions with the true values of the *normalized* features—BWIN_WORLD/seeing and FLUX_MAX/FLUX_APER_{*n*}, where *n* is the aperture size in pixels.

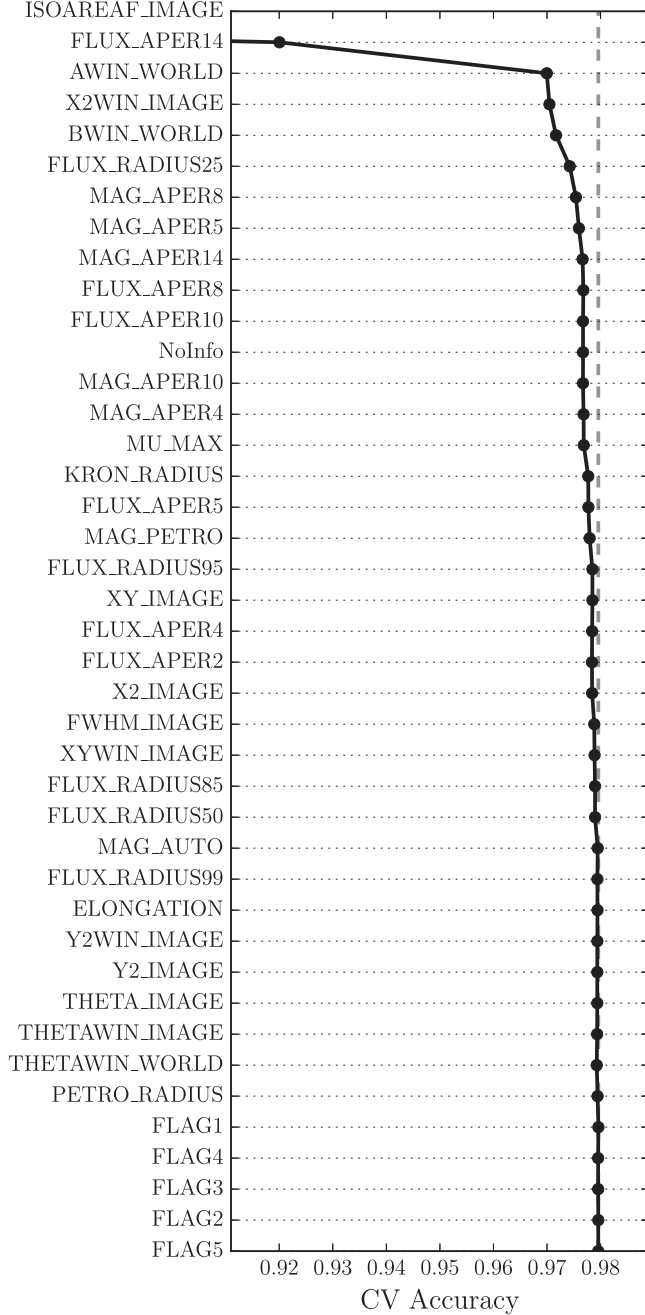


Figure 3. Results from the feature selection procedure. Starting from an empty feature set, features are iteratively added in the order of their RF-ranked importance. The model accuracy progressively increases through the 37th feature, FLAG1. The features listed below FLAG1 are excluded from the final model. The vertical dashed line shows the optimal CV accuracy.

To test the performance of the model on the field, we randomly select 300,000 sources with R.A._{J2000} between 12^h and 13^h and

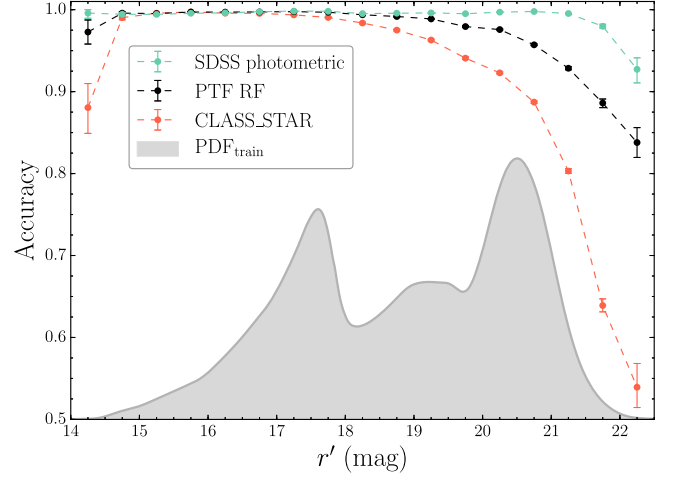


Figure 4. Photometric classification accuracy for SExtractor/CLASS_STAR (pink filled circles), the PTF RF model (black filled circles), and the SDSS photometric pipeline (light-green filled circles) as a function of magnitude for all sources in the test set. A KDE of the PDF of $N(m)$ for training set sources is shown in gray. Accuracies are shown in bins of width 0.5 mag, and the error bars reflect the 95% confidence interval on the mean accuracy from 500 bootstrap resamples in each bin. The PTF RF model shows a significant improvement over CLASS_STAR, especially at faint magnitudes.

decl._{J2000} between 20° and 35° from the PTF reference-image catalogs. This area was selected to test the model at high galactic latitudes ($b > 75^\circ$); we expect blending to be significantly worse near the galactic plane ($|b| \lesssim 15^\circ$), which will in turn degrade the quality of the model. Using SDSS CasJobs, we perform a 1'' cross-match between the randomly selected sources and the SDSS photometric catalog, yielding 280,972 common sources, which we refer to as the SDSS field set. Our RF model predictions produce an overall accuracy of 83.8% when compared to the SDSS photometric classifications, ~15% worse than the ~98% accuracy reported for the test set (Section 4.2). This degradation in performance is expected, as the field population is much fainter than the test set.

Figure 7 shows accuracy and ROC curves to compare the performance of the model on the test set versus the field. The left panel of the figure illustrates that the typical field source is significantly fainter than those present in the training/test set. The middle panel shows that predictions on the test set overstate the accuracy of the model for sources with $r' \gtrsim 20$ mag. This is further corroborated by the right panel, which shows that the test set ROC curve and the ROC curve for bright field sources ($r' \leq 20$ mag) are nearly identical. The ROC curves show successively worse performance when including fainter and fainter sources.

The one caveat to these conclusions is that the SDSS photometric classifications do not truly provide ground truth: Figure 4 shows that the accuracy of the SDSS model drops to

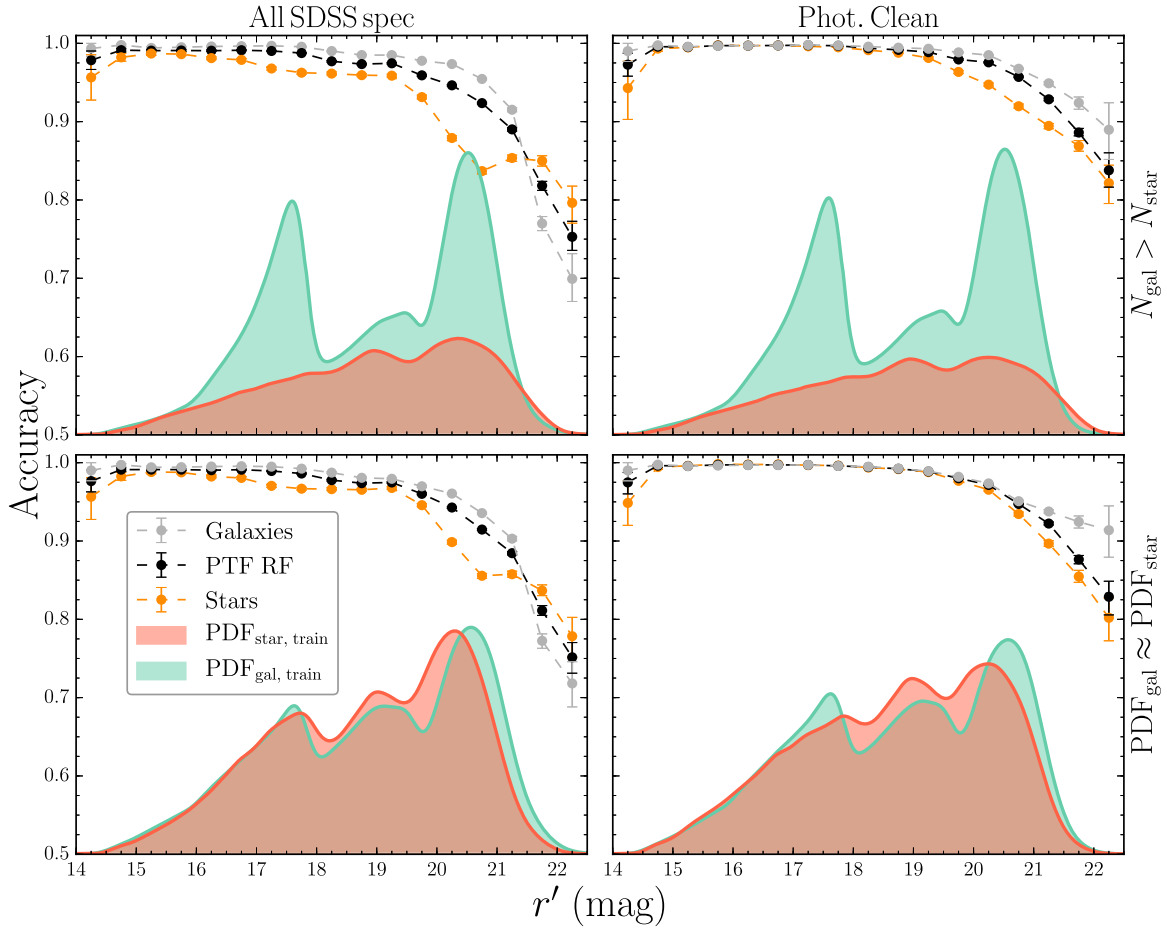


Figure 5. Accuracy of the PTF RF test set predictions as a function of magnitude for different permutations of the training set. In each panel, the black filled circles show the overall accuracy of the model, the gray filled circles show the accuracy when only considering galaxies, and the orange filled circles show the accuracy for stars. Additionally, a KDE of the PDF for stars and galaxies in the training set is shown in pink and light green, respectively. The stellar PDF has been normalized by the ratio of the number of stars to the number of galaxies in the training set. The training set variations are as follows: full training set including all PTF sources with SDSS spectra (upper left), balanced version of the full training set designed to have $\text{PDF}_{\text{gal}} \approx \text{PDF}_{\text{star}}$ (lower left; see text for further details), the *photometrically clean* training set (upper right; see Section 2.3), and the balanced *photometrically clean* training set (lower right). See Figure 4 for a definition of the bin width and uncertainties.

$\sim 93\%$ near the PTF reference-image detection limit. Furthermore, the *photometrically clean* training and test sets overstate the accuracy of all models, as photometric blends have been actively removed. Nevertheless, the results presented in this section are comparative. Sources that are blended in SDSS imaging should also be blended in PTF imaging, meaning that in many of these cases both classifiers will have the same incorrect classification. Thus, the divergence between the two curves in the middle panel of Figure 7 cannot be explained completely by misclassifications by SDSS at the faint end.

The final RF relative ranking used to select a pristine sample of point sources is determined from the ROC curves shown in Figure 7. Prior to selecting an RF relative ranking corresponding to $\text{FPR} = 0.005$, we impose a magnitude cut, $R_{\text{PTF}} \leq 21$ mag, as the accuracy of the SDSS photometric classifier quickly declines for $r' > 21$ mag.¹⁹ There are 463,581,596 PTF sources with $R_{\text{PTF}} \leq 21$ mag, and the $\text{TPR} = 0.695$ at $\text{FPR} = 0.005$ for this subset of the data. This corresponds to an RF relative ranking threshold of 0.966, meaning that only sources classified as point sources in ≥ 725 trees in the forest pass the cut. While as many as $\sim 30\%$ of the *true* $R_{\text{PTF}} \leq 21$ mag point sources are missed by

this cut, our objective is to create a catalog of point sources with virtually no galaxies classified as stars. The vast majority of sources are faint, where classification is the most challenging, meaning that this requirement results in a final point-source catalog that is incomplete. Application of the 0.966 threshold yields a final point-source catalog containing 170,440,636 $R_{\text{PTF}} \leq 21$ mag sources, $\sim 30\%$ of *all* sources extracted from the PTF reference images.

5.2. Comparison to the Previous Star Catalog

Prior to the completion of the RF point-source catalog, the PTF real-time pipeline (Cao et al. 2016) utilized a star catalog based on several cuts on SExtractor parameters. Hereafter, we refer to this initial star catalog as the NERSC catalog. Real-time transient candidates that are spatially coincident with sources in the NERSC catalog are rejected as FPs and removed from the stream prior to human vetting. In the NERSC catalog, stars are defined as all NERSC reference-image sources²⁰

¹⁹ See Figure 4 and http://www.sdss.org/dr12/imaging/other_info/#stargalaxy.

²⁰ The reference images utilized in this study are, on average, slightly deeper than NERSC pipeline references. The same procedure is used to create both sets of references, after which SExtractor is used for source detection.

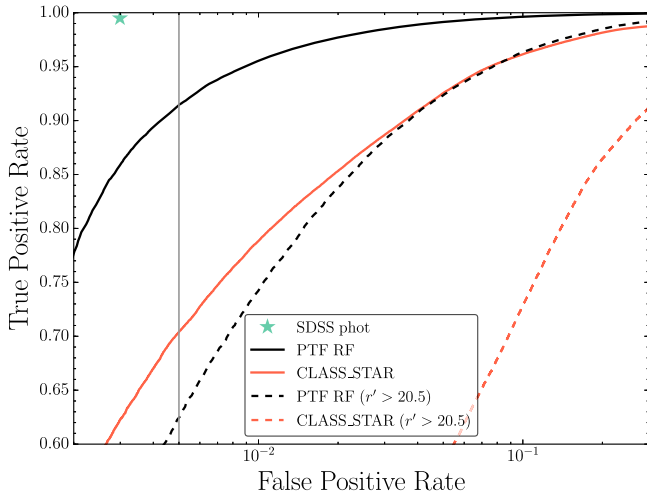


Figure 6. ROC curves comparing the relative performance of SDSS, the PTF RF model, and *SExtractor*. The solid black and red lines show the ROC curve for the PTF RF model and *SExtractor*, respectively, as evaluated by the *photometrically clean* test set. The *SExtractor* ROC curve is generated by varying the classification threshold from `CLASS_STAR` = 0 to 1. The dashed black and red lines show the ROC curves for faint ($r' > 20.5$ mag) *photometrically clean* test set sources for PTF and *SExtractor*, respectively. The solid vertical line shows the desired FPR = 0.005 for the final PTF point-source catalog. The SDSS classifier is shown as a turquoise star, due to the binary nature of the SDSS photometric classification.

satisfying the following cuts:

$$\begin{aligned} \text{MAG_BEST} &< 20, \\ \text{MU_MAX} - \text{MAG_BEST} &< 0.2 \\ &+ \text{med}(\text{MU_MAX} - \text{MAG_BEST}), \\ \text{FWHM_IMAGE} &< 2 \times \text{med}(\text{FWHM_IMAGE}), \\ \text{ELONGATION} &< 1.5, \end{aligned}$$

where `med()` refers to the median value for all sources detected on the same CCD. Initial testing showed that these cuts identified point sources more reliably than `CLASS_STAR`.

To compare the performance of the NERSC catalog and the PTF RF catalog, we adopt the SDSS field set from Section 5.1. Using a $1''$ radial cross-match, there are 241,675 sources in common among the SDSS field set, the NERSC catalog, and the PTF RF catalog. As the NERSC catalog adopts a single hard cut for classification, a comparison of ROC curves is not possible. Instead, we compare the confusion matrix for each, which summarizes the total number of stars classified as stars (TP), galaxies classified as stars (FP), galaxies classified as galaxies (TN), and stars classified as galaxies (FN). Ideally, the confusion matrix would only have power along the diagonal, indicating perfect classification. The PTF RF catalog, however, has been optimized to minimize FP (by adopting a classification threshold > 0.966), resulting in significant off-diagonal power.

Limiting the sample to sources with $R_{\text{PTF}} \leq 21$ mag and adopting the SDSS photometric classifications as ground truth yields the confusion matrices shown in Figure 8. The shading in each matrix shows that the qualitative performance of the catalogs is similar. In detail, however, the PTF RF catalog produces more TPs and, most importantly, a factor of ~ 15 fewer FPs. The NERSC catalog removes $\gtrsim 7.5\%$ of all galaxies from the search for transients. The PTF RF catalog reduces this fraction to $\lesssim 0.5\%$, while also rejecting a larger number of true point sources from the candidate stream. Thus, adoption of the

new PTF RF catalog significantly improves the search for all transients relative to the NERSC catalog.

5.3. Demonstrable Improvements in the Discovery Potential of the PTF RF Catalog

While Section 5.2 provides evidence that $\sim 7\%$ of galaxies are misclassified as stars by the NERSC catalog, here we provide definitive examples of transients PTF missed that would have been detected had the PTF RF catalog been employed. These transients were identified via a *non-exhaustive* search, which included the following steps:

1. All transient candidates with the NERSC flag `is_star` and satisfying the normal thresholds for human vetting during the period from 2015 November 01.0 UT to 2016 January 01.0 UT were selected. This selection yielded 72,546 unique sources that were detected between 2 and 170 times during the search period.
2. These $\sim 72\text{k}$ sources were cross-matched against the PTF RF catalog to identify candidates classified as stars in the NERSC catalog and galaxies in the PTF RF catalog, resulting in a list of 25,138 sources.
3. Those candidates with detections before 2015 August 01.0 UT or after 2016 March 01.0 UT were removed to exclude long-term variables, which reduced the list to 15,737 candidates.
4. These were cross-matched to SDSS to provide color and morphological information, further culling the list to 7813 sources.
5. The 39 sources classified as galaxies by SDSS with ≥ 5 detections between 2015 November 01.0 UT and 2016 January 01.0 UT and $u'_0 - g'_0 > 0.7$ mag were visually inspected. The color cut was applied to eliminate likely QSOs (see, e.g., Figure 4 in Sesar et al. 2007).

Visual inspection of these sources revealed two transient candidates that were otherwise missed by the NERSC discovery pipeline. These candidates have been internally designated as iPTF 15eyh²¹ and iPTF 16cbx, and their light curves are shown in Figure 9. The lack of historical variability and host galaxy colors suggest that these candidates are bona fide transients and not active galactic nuclei. The nature of these transients is difficult to discern given the partial light-curve coverage and lack of spectroscopic observations. Nevertheless, our limited, nonexhaustive search reveals that PTF missed several transients as a result of misclassifications in the NERSC catalog. The use of the PTF RF catalog will significantly lower the number of transients that are missed because their host galaxies are classified as stars.

5.4. Supplementing the Catalog with SDSS

It is possible to further improve the rejection of point sources from the candidate transient stream using SDSS imaging data, which is limited to approximately half the total PTF imaging footprint. As previously discussed, SDSS has superior imaging quality to PTF and provides superior photometric classifications (see Figure 4). The unfiltered addition of SDSS stars to

²¹ PTF transients are named based on the year when they are discovered, i.e., when a human manually saves a candidate as real. In late 2015, the PTF IPAC pipeline (see Masci et al. 2016) used a preliminary version of the PTF RF catalog to reject stars, and as a result, iPTF 15eyh was successfully identified in real time. We include it here because the NERSC pipeline missed this transient.

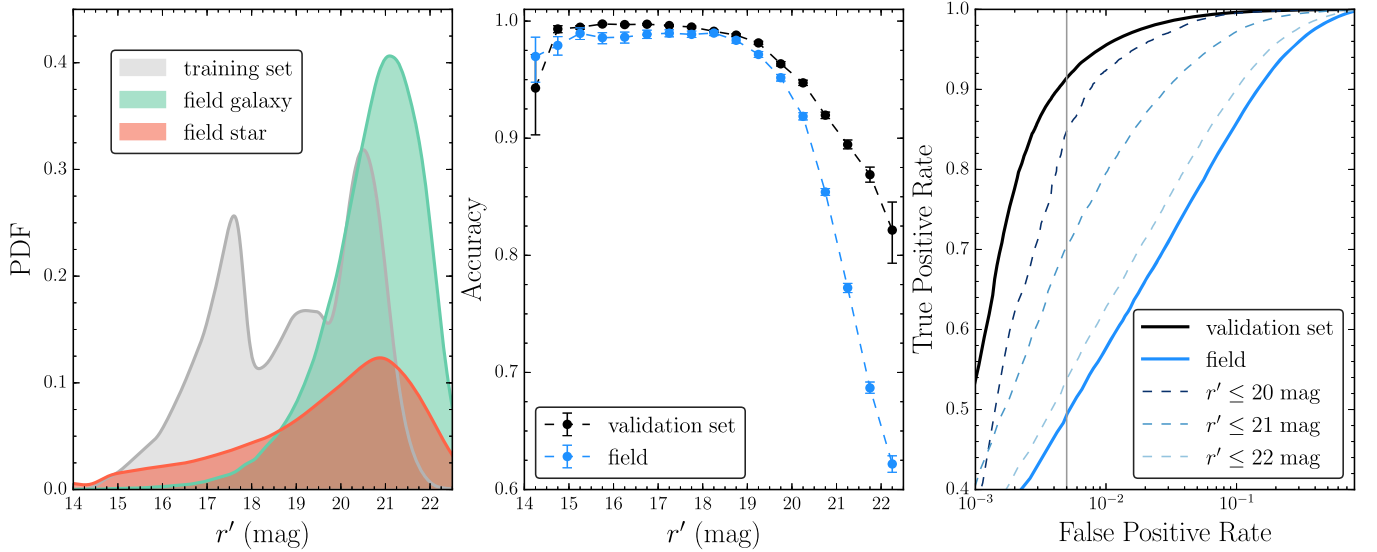


Figure 7. Accuracy and ROC curves for the PTF RF model compared to SDSS photometric classifications. Left: KDEs of the magnitude PDF for the full training set (gray), SDSS photometric stars (pink), and SDSS photometric galaxies (light green). Middle: accuracy curves for the PTF RF model as tested on the test set (black filled circles) and tested by the SDSS photometric classification (light-blue filled circles). The performance on the random set of field sources shows that the test set predictions overstate the true accuracy of the model. Right: ROC curve for the test set predictions (solid black line) and the predictions for SDSS field sources (solid blue line). The dashed lines show ROC curves when constraining the field sample to stars brighter than 20, 21, and 22 mag. The solid vertical line shows the FPR = 0.005 cut adopted for inclusion in the final PTF point-source catalog.

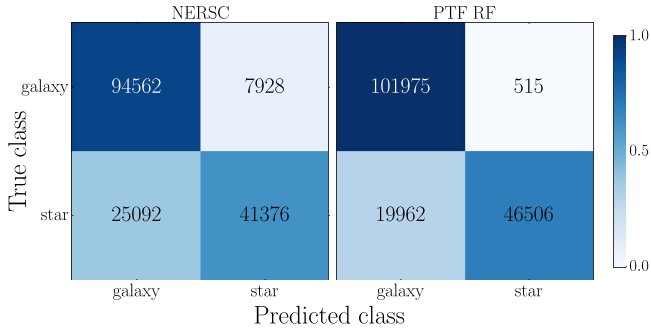


Figure 8. Confusion matrix comparison between the NERSC catalog and the PTF RF catalog. Each matrix shows (clockwise, from the upper left) the total TN, FP, TP, and FN. The colors represent the fraction of true class members. The PTF RF catalog has a $\sim 12\%$ improvement in TPR and, more importantly, a factor of ~ 15 decrease in the FPR relative to the NERSC catalog.

the PTF point-source catalog will reduce its effectiveness, however, as the SDSS classification has not been tuned to produce an FPR = 0.005. Below, we adjust the SDSS classification threshold to produce the desired FPR.

The SDSS pipeline classifies a source as a star if

$$f_{\text{PSF}}/f_{\text{cmodel}} > 0.875,$$

where f_{PSF} is the PSF flux and f_{cmodel} is the composite-model flux, which measures the best-fit linear combination of an exponential and a de Vaucouleurs profile. The final classification is performed using the sum of the fluxes in all bands where the source is detected. Adjusting the decision threshold up or down decreases or increases FP, respectively.

To determine the optimal threshold for $f_{\text{PSF}}/f_{\text{cmodel}}$, we select spectroscopic classifications and photometric fluxes for all SDSS sources. The query is performed via *CasJobs* to select sources with *sciencePrimary* = 1 and *mode* = 1, corresponding to the primary spectroscopic and photometric detection of a given source. A total of 3,537,411 sources match this criteria, which we hereafter refer to as the SDSS specphot sample. We perform an

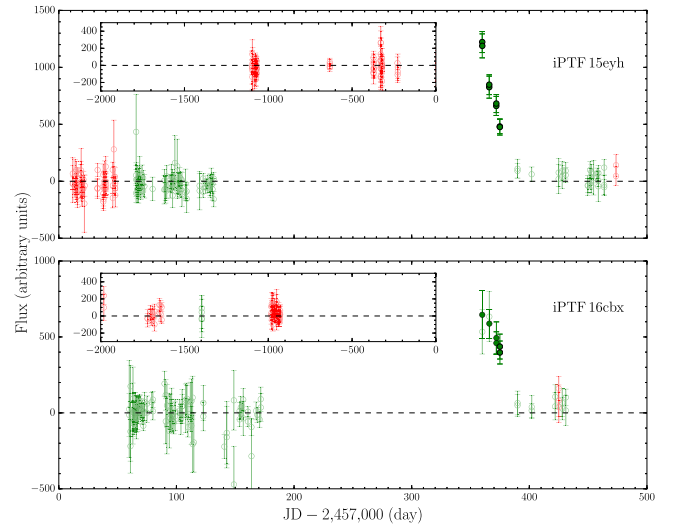


Figure 9. Light curves for the two transients missed by the NERSC pipeline, produced via difference-image-PSF photometry at the location of the transient. PSF flux measurements are shown with arbitrary units. The *g*-band observations are shown in green, while *R*-band observations are shown in red. Epochs where the transient is detected, i.e., signal-to-noise ratio ≥ 4 , are shown with filled circles, while epochs with no detection are shown with open circles. Both transients were detected over a ~ 2 -week period starting on ~ 2015 December 01 UT. The inset panels show the lack of historical variability over the duration of PTF observations.

ROC-like analysis to measure changes in the TPR and FPR as a function of classification threshold, where we have adjusted $f_{\text{PSF}}/f_{\text{cmodel}}$ from its highest value to its lowest. The results of this procedure are shown via the solid black line in Figure 10. The vertical gray line shows the desired FPR = 0.005, while the filled black star shows the location on the curve corresponding to $f_{\text{PSF}}/f_{\text{cmodel}} = 0.875$. Thus, adopting the SDSS classification threshold would yield a TPR ≈ 0.93 and FPR ≈ 0.02 , which is significantly higher than our target. It is clear that an alternative threshold is needed to achieve an FPR = 0.005.

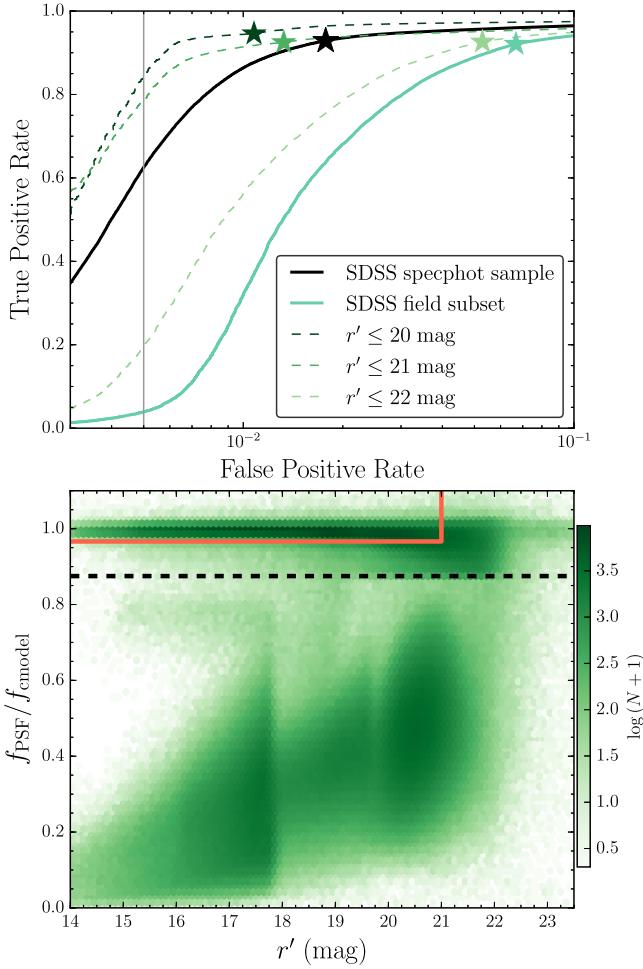


Figure 10. Decision thresholds for selecting SDSS point sources with $\text{FPR} = 0.005$. Top: ROC-like curves (see text) for SDSS photometric classification. For each curve the filled star marks the location corresponding to $f_{\text{PSF}}/f_{\text{cmodel}} = 0.875$, the SDSS pipeline classification threshold. The solid black and light-green lines show the SDSS specphot sample and field subset, respectively. The dashed lines show the ROC-like curves for the SDSS field subset restricted to sources brighter than $r' = 20, 21$, and 22 mag from top to bottom. For our purposes, the 0.875 threshold produces too many misclassified galaxies. Bottom: Density plot showing $f_{\text{PSF}}/f_{\text{cmodel}}$ as a function of r' for all sources in the SDSS specphot sample. Pixels are ~ 0.1 mag wide. The concentration at $f_{\text{PSF}}/f_{\text{cmodel}} \approx 1$ corresponds to point sources. The dashed horizontal line represents the SDSS classification threshold, while we only classify those sources enclosed by the solid pink line as point sources. Our cut is more restrictive and produces a factor of ~ 3 decrease in the number of galaxies erroneously classified as point sources.

The SDSS specphot sample is heavily biased by the SDSS spectroscopic-targeting function and as such does not reflect the true distribution of SDSS photometric detections. This is illustrated by the gray distribution in the left panel of Figure 7,²² as compared to the pink and light-green distributions. The optimal SDSS classification threshold should be selected from a set of sources that reflect the true distributions found in nature. We approximate such a set of sources via a weighted random subset of the SDSS specphot sample. The individual weights are determined via KDEs of the magnitude PDFs for the photometric sample and the specphot sample. The PDFs are evaluated at the r' mag of each source, with the individual weights equal to the

photometric sample PDF divided by the specphot PDF. These weights emphasize faint sources, which are underrepresented in the SDSS specphot sample. As galaxies outnumber stars by a ratio of $\sim 2:1$ in the SDSS photometric observations, a weighted random selection of 200,000 galaxies and 100,000 stars from the SDSS specphot sample is made. We hereafter refer to these 300,000 sources as the SDSS field subset.

The ROC-like curve for the SDSS field subset is shown via the solid light-green line in Figure 10. Again, the filled star shows the location of the $f_{\text{PSF}}/f_{\text{cmodel}} = 0.875$ threshold. Adopting the SDSS classification for all photometric sources detected by SDSS would yield an $\text{FPR} \approx 0.07$. Additionally, a requirement of $\text{FPR} = 0.005$ over all SDSS photometric detections would yield a $\text{TPR} \approx 0.04$, which is so small that it is effectively useless for screening point sources from the transient-candidate stream. Figure 10 also shows the ROC-like curves for the SDSS field subset restricted to sources with $r' \leq 20, 21$, and 22 mag via dashed lines from top to bottom. The dashed lines confirm the previous assertions that the fidelity of the SDSS photometric classifier degrades rapidly for $r' > 21$ mag. Thus, we supplement the PTF point-source catalog with all SDSS photometric detections satisfying $r' \leq 21$ mag and $f_{\text{PSF}}/f_{\text{cmodel}} > 0.9658$.²³ For sources with $r' \leq 21$ mag this corresponds to a $\text{TPR} = 0.79$ at the desired $\text{FPR} = 0.005$. Thus, relative to the PTF RF point-source catalog, SDSS provides a $\sim 12\%$ increase in the recovered point sources at the desired FPR .

The difference between our selection of SDSS point sources and that of the SDSS pipeline is illustrated in the bottom panel of Figure 10. The density of the SDSS specphot sample is shown in the $f_{\text{PSF}}/f_{\text{cmodel}} - r'$ plane. There is a clear delineation between sources with $f_{\text{PSF}}/f_{\text{cmodel}} \approx 1$ and those with a larger model flux than PSF flux. The horizontal dashed line represents the SDSS classification threshold, while we only classify those sources enclosed by the solid pink line as point sources. Our cut is more restrictive and produces a factor of ~ 3 decrease in the number of galaxies erroneously classified as point sources.

6. SUMMARY AND CONCLUSIONS

We have presented a method for the automated classification of stars and galaxies in PTF imaging data. The classifier utilizes the RF algorithm and is trained using $> 3 \times 10^6$ PTF sources with SDSS spectra. A non-negligible fraction of point sources in the training set ($\sim 2\%$) are photometric blends, targeted owing to the SDSS bias to observe galaxies, especially LRGs. These blends, along with compact galaxies and low-redshift quasars, were removed from the training set to improve the overall performance of the classifier. Features were selected from SExtractor shape and brightness measurements, and the model tuning parameters were optimized via CV on the training set.

We showed that the final PTF RF model outperformed the SExtractor classifications, with an overall improvement of $\sim 4\%$ on the *photometrically clean* test set and a more impressive $\sim 19\%$ improvement for sources with $r' \geq 21$ mag. Within the SDSS footprint, which covers roughly half

²² Strictly speaking, Figure 7 shows the distribution of spectroscopic training sources for the PTF RF model, which is virtually indistinguishable from the SDSS specphot sample.

²³ The online SDSS documentation states that sources with $m_{\text{PSF}} - m_{\text{cmodel}} < 0.145$ mag are classified as stars, which is equivalent to the $f_{\text{PSF}}/f_{\text{cmodel}} > 0.875$ threshold discussed here. In terms of magnitude difference, the adopted point-source classification threshold corresponds to $m_{\text{PSF}} - m_{\text{cmodel}} < 0.037$ mag.

of the total PTF imaging area, the SDSS pipeline provides better classifications than the PTF RF model, due to the superior seeing in SDSS images. The PTF RF model produces near-perfect separation of stars and galaxies down to ~ 19 mag. Tests on a random selection of field stars show that the classification accuracy remains above 80% down to ~ 21 mag. To generate our final PTF point-source catalog, we apply a conservative classification cut, designed to produce an $\text{FPR} = 0.005$. Ultimately, only sources classified as stars in ≥ 725 of the 750 RF trees, corresponding to an RF relative ranking of > 0.966 , are included.

In sum, there are $\sim 1.70 \times 10^8$ sources in the point-source catalog, of which only $\sim 10^6$ are expected to be galaxies. Following a *nonexhaustive* search for transients missed by the NERSC catalog, we identify two transients that would have been detected using the PTF RF catalog. This search, which only covered the last 2 months of 2015, provides definitive evidence that the PTF RF catalog enables new discoveries. We have additionally developed a new method to select SDSS point sources with an $\text{FPR} = 0.005$, which we use to supplement the PTF RF point-source catalog within the SDSS imaging footprint. The inclusion of these SDSS sources increases the number of point-source detections by $\sim 12\%$. The catalog has been incorporated into the various PTF transient discovery pipelines, and candidates associated with point sources are now automatically rejected and removed from the stream.

Despite the large number of sources in the PTF point-source catalog, our conservative cut on RF relative ranking means that $\sim 8 \times 10^7$ point sources are currently excluded from the catalog (more if one includes the sources fainter than 21 mag). Moving forward, especially with an eye toward ZTF, there are several potential improvements that could be made to improve the fidelity of the model, particularly at faint magnitudes.

PTF, which has been running since 2009, uses `SExtractor` version 2.8.6 in the IPAC imaging pipelines. Recent versions of `SExtractor` (e.g., v2.19.5) include a new parameter, `SPREAD_MODEL` (Desai et al. 2012), which acts as a discriminant between the best-fit PSF model and an exponential model. Initial tests with `SPREAD_MODEL` show that it is useful for separating stars and galaxies (Soumagnac et al. 2015). The inclusion of `SPREAD_MODEL` in a ZTF star-galaxy model will yield improvements relative to the PTF RF model.

Additional improvements can be had via deeper co-adds, which will make it easier to detect extended emission from sources with $r' \gtrsim 22$ mag. ZTF surveys the sky at a rate that is $\sim 15\times$ faster than PTF, which will enable the creation of deep reference images faster than is currently possible. For instance, $\sim 38\%$ of the sources in the training set were taken from co-adds of only five images, while $\sim 57\%$ are from co-adds of ≤ 10 images. These references have a depth similar to SDSS, while the co-adds of 50 images detect sources as faint as $R_{\text{PTF}} \gtrsim 23.5$ mag. For ZTF deeper reference images will be generated to construct a point-source catalog with higher fidelity.

Finally, altogether superior modeling of the sources at the image level could improve the separation of stars and galaxies in PTF and ZTF data. This could include techniques as familiar as simple PSF fitting with `DAOPHOT` (Stetson 1987), to more

advanced solutions that construct probabilistic models of the data, such as the `Tractor`²⁴ (Lang et al. 2016).

The detection and characterization of fast transients in the coming years will be as much about software development as it is about improvements in instrumentation. While events that evolve and disappear on timescales $\lesssim 24$ hr have already been discovered (e.g., Cenko et al. 2013), future \sim real-time classifications of these rarities will require swift automated decisions. The optimal allocation of expensive follow-up resources requires the best possible rejection of FPs. One step in that direction is to identify as many faint stars as possible, as we have done here for PTF observations. While the primary motivation for constructing the PTF RF catalog is to better enable the search for fast transients, these efforts ultimately improve the search for all extragalactic transients.

We are now firmly in the age of GW detections (Abbott et al. 2016b), and the identification of an electromagnetic counterpart to a GW event stands out as one of the most challenging problems in astrophysics in the coming years. The search for such counterparts will monopolize the use of wide-field telescopes across the globe (e.g., Abbott et al. 2016a). Without some means to significantly reduce the haystacks, however, the search for these needles will be hopeless. Minimizing the stages at which human inspection and intervention are required, by actively reducing the number of FP candidates, will improve our chances of one day catching the elusive transients associated with GW events.

This project started as part of an undergraduate research project at the California Institute of Technology. We thank T. Prince for funding M.K.K. during the summer of 2015.

We are extremely grateful to A. Thakar and the entire SDSS CasJobs Helpdesk for assistance in performing the large cross-match between PTF and SDSS spectroscopic sources. Without their assistance this study would not have been possible. Without the patience and aid of R. Lupton we would not have been able to recreate the SDSS photometric classifier. We are in debt to M. M. Kasliwal, who endured countless conversations on the appropriate threshold for point-source classification. With gratitude, we salute S. B. Cenko for useful suggestions on the comparison of the NERSC catalog and the PTF RF catalog. Finally, we thank the anonymous referee for suggestions that improved this manuscript.

A.A.M. acknowledges support for this work by NASA from Hubble Fellowship grant HST-HF-51325.01, awarded by STScI, operated by AURA, Inc., for NASA, under contract NAS 5-26555. Part of the research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III Web site is <http://www.sdss3.org/>.

Facilities: Sloan, PO:1.2m.

REFERENCES

- Abazajian, K., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2003, *AJ*, **126**, 2081
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, *ApJL*, **826**, L13
- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, *PhRvL*, **116**, 061102

²⁴ See <http://thetractor.org/>.

- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, **219**, 12
- Anderson, L., Aubourg, É., Bailey, S., et al. 2014, *MNRAS*, **441**, 24
- Ball, N. M., Brunner, R. J., Myers, A. D., & Tcheng, D. 2006, *ApJ*, **650**, 497
- Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2006, *ApJL*, **642**, L137
- Belokurov, V., Zucker, D. B., Evans, N. W., et al. 2007, *ApJ*, **654**, 897
- Berger, E., Chornock, R., Lunnan, R., et al. 2012, *ApJL*, **755**, L29
- Berger, E., Leibler, C. N., Chornock, R., et al. 2013, *ApJ*, **779**, 18
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bloom, J. S., Richards, J. W., Nugent, P. E., et al. 2012, *PASP*, **124**, 1175
- Breiman, L. 1996, *Machine Learning*, 24, 123
- Breiman, L. 2001, *Machine Learning*, 45, 5
- Brink, H., Richards, J. W., Poznanski, D., et al. 2013, *MNRAS*, **435**, 1047
- Cao, Y., Nugent, P. E., & Kasliwal, M. M. 2016, *PASP*, **128**, 114502
- Cenko, S. B., Kulkarni, S. R., Horesh, A., et al. 2013, *ApJ*, **769**, 130
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- Desai, S., Armstrong, R., Mohr, J. J., et al. 2012, *ApJ*, **757**, 83
- Dubath, P., Rimoldini, L., Süveges, M., et al. 2011, *MNRAS*, **414**, 2602
- Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, *AJ*, **122**, 2267
- Hastie, T., Tibshirani, R., & Friedman, J. 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.; Berlin: Springer)
- Ho, S., Agarwal, N., Myers, A. D., et al. 2015, *JCAP*, **5**, 040
- Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, *ApJ*, **673**, 864
- Kasen, D., Fernández, R., & Metzger, B. D. 2015, *MNRAS*, **450**, 1777
- Kasliwal, M. M., Cenko, S. B., Singer, L. P., et al. 2016, *ApJL*, **824**, L24
- Kulkarni, S. R. 2012, in IAU Symp. 285, *New Horizons in Time Domain Astronomy*, ed. E. Griffin, R. Hanisch, & R. Seaman (Cambridge: Cambridge Univ. Press), 55
- Kulkarni, S. R. 2013, *ATel*, **4807**, 1
- Kulkarni, S. R., & Rau, A. 2006, *ApJL*, **644**, L63
- Laher, R. R., Surace, J., Grillmair, C. J., et al. 2014, *PASP*, **126**, 674
- Lang, D., Hogg, D. W., & Schlegel, D. J. 2016, *AJ*, **151**, 36
- Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, *PASP*, **121**, 1395
- Lupton, R. H., Ivezić, Z., Gunn, J. E., et al. 2002, *Proc. SPIE*, **4836**, 350
- Masci, F., Laher, R., Rebbapragada, U., et al. 2017, *PASP*, **129**, 014002
- Metzger, B. D., & Berger, E. 2012, *ApJ*, **746**, 48
- Miller, A. A., Bloom, J. S., Richards, J. W., et al. 2015, *ApJ*, **798**, 122
- Nissanke, S., Kasliwal, M., & Georgieva, A. 2013, *ApJ*, **767**, 124
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
- Rau, A., Kulkarni, S. R., Law, N. M., et al. 2009, *PASP*, **121**, 1334
- Rebbapragada, U., Bue, B., & Wozniak, P. R. 2015, in AAS Meeting 225 Abstracts, **#434.02**
- Richards, J. W., Starr, D. L., Butler, N. R., et al. 2011, *ApJ*, **733**, 10
- Richards, J. W., Starr, D. L., Miller, A. A., et al. 2012, *ApJS*, **203**, 32
- Ross, A. J., Ho, S., Cuesta, A. J., et al. 2011, *MNRAS*, **417**, 1350
- Scott, D. W. 1992, *Multivariate Density Estimation* (New York: Wiley)
- Sesar, B., Ivezić, Ž., Lupton, R. H., et al. 2007, *AJ*, **134**, 2236
- Soumagnac, M. T., Abdalla, F. B., Lahav, O., et al. 2015, *MNRAS*, **450**, 666
- Stekhoven, D. J., & Bühlmann, P. 2012, *Bioinformatics*, **28**, 112
- Stetson, P. B. 1987, *PASP*, **99**, 191
- Vasconcellos, E. C., de Carvalho, R. R., Gal, R. R., et al. 2011, *AJ*, **141**, 189
- Yasuda, N., Fukugita, M., Narayanan, V. K., et al. 2001, *AJ*, **122**, 1104
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579